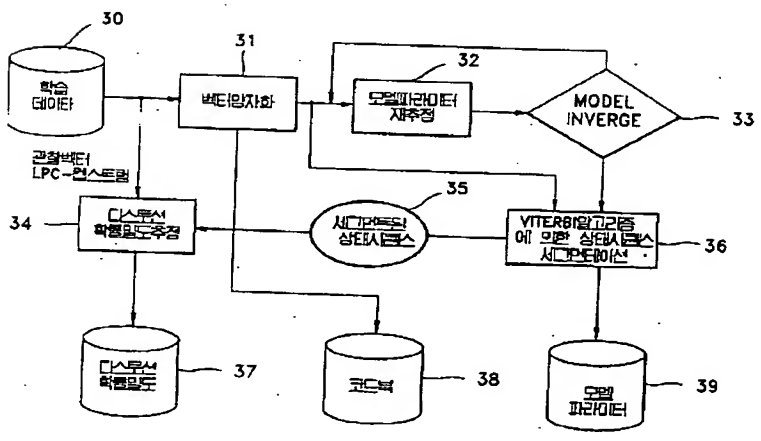
KOREAN PATENT ABSTRACTS(KR)

Document Code:B1
(11) Publication No.1001703170000 (44) Publication.Date. 19981014
(21) Application No.1019950020651 (22) Application Date. 19950713
(51) IPC Code:
G10L 9/14

(71) Applicant:
SAMSUNG ELECTRONICS CO., LTD.
(72) Inventor:
YANG, JUN YONG
(30) Priority:

(54) Title of Invention
METHOD FOR RECOGNIZING SPEECH BY USING HIDDEN MARKOV MODEL HAVING DISTORTION DENSITY OF OBSERVATION VECTOR

Representative drawing



(57) Abstract:

PURPOSE: A method for recognizing speech by using a hidden Markov model having distortion density of an observation vector is provided to consider time structure of observation sequence and characteristics of LPC-cepstrum.

CONSTITUTION: A codebook is made by quantizing each vector from training data(31). The vector-quantized training data is repeatedly re-estimated until a model parameter reaches appropriate level(32). A state sequence is segmented by Viterbi algorithm for generating model parameter(36). Distortion density of training data is estimated based on the segmented state sequence (34,37). If data to be recognized is inputted by using the codebook, the model parameter

and the distortion density, the input data are transformed into an observation sequence by using the codebook, so that likelihood obtained by Viterbi algorithm, segmented state sequence generated by the Viterbi procedure, and distortion density obtained from distortion of an input observation vector are operated for recognizing speech.

(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) Int. Cl.[6]
G10L 9/14

(45) 공고일자 1999년03월30일
(11) 등록번호 특0170317
(24) 등록일자 1998년10월14일

(21) 출원번호 특1995-020651
(22) 출원일자 1995년07월13일
(65) 공개번호 특1997-007791
(43) 공개일자 1997년02월21일

(73) 특허권자 삼성전자주식회사 김광호
경기도 수원시 팔달구 매탄동 416번지
(72) 발명자 양준용
서울특별시 관악구 봉천2동 41-183
(74) 대리인 이영필, 권석흥, 오규환

심사관 : 박정학

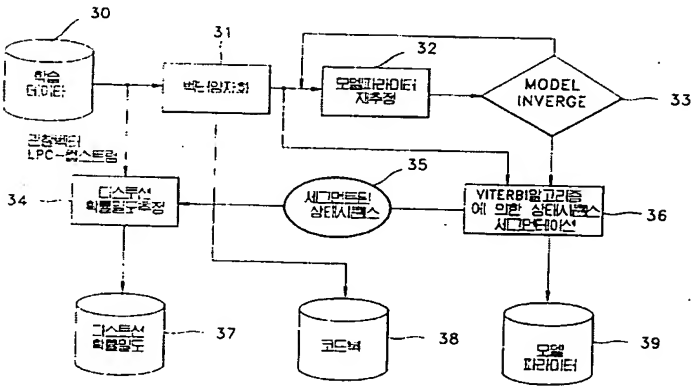
(54) 관찰벡터의 디스토션 확률밀도를 가진 은닉마코프 모델을 이용한 음성인식 방법

요약

본 발명은 관찰벡터의 디스토션(Distortion)을 이용한 음성인식에 관한 것으로서, 관찰 벡터의 디스토션(Distortion) 확률밀도(density)를 가진 은닉 마코프 모델(HMM)의 학습과정시 디스토션(Distortion)확률밀도(density)를 구하는 방법은 상태 세그먼테이션 시퀀스를 구하는 단계:, 상기 상태 세그멘테이션된 학습관찰 시퀀스를 참조하여 각 상태 i의 평균길이 mi를 구하는 단계:, 상기 상태 i에서의 정규화 시간 구간 j 마다 평균 벡터 v_{ij}를 구하는 단계:, 각각의 학습관찰 시퀀스 k에 대해 각 상태 i에서의 정규화 시간 구간 j 마다 평균벡터에 대한 디스토션 eijk를 구하는 단계: 및 각 상태 i에서의 정규화 시간구간 j 마다 총 k개의 학습 관찰 벡터들의 디스토션 중 최소값(n_{ij})과 최대값(x_{ij})를 구하는 단계를 포함한다.

상술한 바와 같이 본 발명은 일반적인 HMM에 비해 나은 음성 인식률을 제공한다.

대표도



명세서

[발명의 명칭]

관찰벡터의 디스토션 확률밀도(Distortion density)를 가진 은닉 마코프 모델(Hidden Markov Model)을 이용한 음성인식 방법

[도면의 간단한 설명]

제1도는 디스토션 확률밀도를 구하기 위한 단계를 설명하기 위한 도로써, 10단계, 12단계, 14단계를 설명하기 위한 도이다.

제2도는 디스토션 확률밀도를 구하기 위한 단계를 설명하기 위한 도로써, 16단계, 18단계, 20단계를 설명하기 위한 도이다.

제3도는 본 발명에서 구현하는 HMM/Distortion 인식기의 학습과정의 블럭 다이어그램이다.

제4도는 HMM/Distortion 인식기의 인식과정의 블럭 다이어그램이다.

[발명의 상세한 설명]

본 발명은 음성인식에 관한 것으로서, 더욱 상세하게는 관찰벡터의 디스토션(Distortion)을 이용한 음성인식에 관한 것이다.

최근에는 동적 시간축 정합(dynamic time warping)기법(DTW), 벡터 양자화(vector qauntization; 이하 VQ)기법, 및 은닉 마코프 모델(Hidden Markov Model; 이하 HMM)등이 음성인식 응용분야에 적용되고 있다.

이 중에서도 상기 상기 은닉 마코프 모델(HMM)은 여러개의 학습패턴을 확률적으로 모델화하는 방법으로서, 화자내의 음성변이를 흡수하면서도 다수의 화자(speaker independent)를 대상으로 높은 인식률과 빠른처리가 가능한 장점이 있다. 이 HMM의 주된 특징 중의 하나는 음성신호의 시간적인 구조(temporal structure)를 내재적(implicit)으로나 명료(explicit)하게 모델화하려는 시도로서 나타난다. 이러한 시간적인 구조에 의해서 음성신호에 나타나는 음향이벤트(acoustic event)의 시간적인 지속시간(duration)이나 상대적인 순서에 어떠한 의미를 부여할 수 있다. 이러한 상태 지속(state duration)의 모델링은 일반적인 은닉 마코프 모델(HMM)에 결합되어 보다 좋은 인식률을 보이고 있다.

그러나, 상태 지속을 명백하게 모델링할 때는 학습 데이타의 양이 충분해야 올바른 결과를 얻을 수 있을 뿐 아니라, 특히 비-매개변수(non-parametric)기법의 경우에는 모든 상태 지속과 상태에 대하여 확률밀도 함수를 구하기 때문에 계산이 복잡해져서 실용적이지 못하고, 매개변수(parametric)기법의 경우에는 학습데이타가 적은 경우에는 포아송(poison), 감마(gamma) 등 대응하는 확률밀도 함수식으로부터 상태지속 확률함수를 추정해야 인식률의 저하를 막을 수 있다. 그러나 이렇게 확률밀도 함수식을 이용하는 경우에는, 인식하고자 하는 모든 단어들의 모든 상태에 대하여확률밀도 함수식이 필수적으로 적절할 필요가 없다는 가정이 잘 적용될 때에만 가능하다는 문제점이 있다.

따라서, 본 발명은 음성신호의 시간적인 구조와, 관찰벡터의 특징을 고려할 수 있는 관찰벡터의 디스토션 밀도(distortion density)를 가진 은닉 마코프모델(HMM)(이하, HMM/Distortion)을 이용한 음성인식방법을 제공하는 것을 그 목적으로 한다.

상기의 목적을 달성하기 위한 음성인식방법은 학습데이타를 백터양자화시켜 코드북을 만들고, 상기 벡터양자화된 학습데이타의 모델파라메터를 재추정한 후, 비터비 알고리즘에 의해 상태시퀀스를 세그먼트하여 모델파라메터를 생성하고, 상기 세그먼트된 상태시퀀스를 근거로 상기 학습데이타의 디스토션 확률밀도를 추정하는 학습과정과, 상기 학습과정에서 생성된 코드북과 모델파라미터 및 디스토션 확률 밀도를 이용하여, 인식하고자 하는 데이타가 입력되면, 상기 코드북을 이용하여 상기 입력데이타를 관찰시퀀스로 변환시켜, 비터비 알고리즘에 의해 구해진 확률($\rho(o \setminus \lambda)$)과, 상기 비터비 절차에 의해 생성된 세그먼트된 상태시퀀스와 입력 관찰 벡터(LPC-켑스트럼)의 디스토션으로부터 구해진 디스토션 확률을 연산하여 음성을 인식하는 과정을 포함하는 것을 특징으로 한다.

이하, 첨부된 도면을 참조하여 본 발명을 보다 상세히 설명한다.

제3도는 본 발명의 HMM/Distortion 인식기에서 학급과정을 설명하기 위한 도면이고, 제4도는 그에 따른 인식과정을 설명하기 위한 도면이다.

제3도에 도시된 학습과정에서는 학습데이타를 벡터양자화시켜 코드북을 만들고, 상기 벡터양자화된 학습데이타를 모델파라메터를 재추정한 후, 비터비 알고리즘에 의해 상태시퀀스를 세그먼트하여 모델파라메터를 생성하고, 상기 세그먼트된 상태시퀀스를 근거로 상기 학습데이타의 디스토션 확률밀도를 추정한다.

제4도에 도시된 인식과정에서는 제3도에 도시된 학습과정에서 생성된 코드북과 모델파라미터 및 디스토션 확률밀도를 이용하여, 인식하고자 하는 입력데이타가 입력되면, 상기 코드북을 이용하여 상기 입력데이타를 관찰시퀀스로 변환시켜, 비터비 알고리즘에 의해 구해진 확률($\rho(o \setminus \lambda)$)과, 상기 비터비 절차에 의해 생성된 세그먼트된 상태시퀀스와 입력 관찰 벡터(LPC-켑스트럼)의 디스토션으로부터 구해진 디스토션 확률을 연산하여 음성을 인식한다.

제3도에 학습과정에서는 HMM/Distortion을 이용하여 관찰 시퀀스들을 학습한 후, 인식과정에서 사용되기 위해 저장되는 파라미터는 일반적인 HMM의 상태천이 확률분포, 관찰 심볼 확률분포, 초기상태 확률분포와 학습관찰 시퀀스들을 벡터 양자화하여 구성한 코드북을 생성한다. 즉, 학습데이타(38)로부터 각 벡터의 양자화(VQ) 단계(31)를 통해 코드북(38)이 만들어지고, 상기 코드북(30)은 인식시에 입력벡터를 코드북 인덱스로 매핑하는데 사용되게 된다. 또한 상기 벡터 양자화 단계(31)로부터 모델의 파라미터(32)가 적정 수준에 도달할 때까지 재추정 과정(33)을 반복한다. 또한 상기 벡터 양자화 단계(31)에 의해 코드북 인덱스로 변환된 학습관찰 시퀀스 집합을 재추정하는 과정으로 얻어진 모델(λ)을 근거로 하여 상태별로 세그먼테이션하는 상태 세그먼테이션 단계(36)는 Viterbi 알고리즘 절차에 의하여 최적의 상태 시퀀스를 구한 후 백트래킹하여 최적의 경로를 구한다.

상기 세그먼테이션 단계(36) 후 모델 파라미터(39)를 형성하고, 이 시퀀스된 상태 시퀀스(35)를 근거로 하여 학습데이타(30)의 관찰 벡터(즉, LPC-켑스트럼)의 거리 확률밀도를 추정하여(34), 거리 확률밀도를 생성(37)시키게 된다.

여기서, 제1도와 제2도에 도시된 디스토션 확률밀도(density)를 구하는 방법은 상태 세그멘테이션 시퀀스를 구하는 단계(10단계)와, 전체 관찰시퀀스에 대한 각 상태(i)의 평균길이(m_i)를 구하는 단계(12단계)와, 각 상태(i)에서의 정규화 시간구간(j) 마다 평균벡터(v_{ij}))를 구하는 단계(14단계); 각각의 학습 관찰 시퀀스(k)에 대해 각 상태(i)에서의 정규화 시간 구간(j) 마다 평균벡터에 대한 디스토션(ejk)를 구하는 단계(16단계)와, 각 상태(i)에서의 정규화 시간구간(j)마다 총 k개의 학습 관찰 벡터들의 디스토션 중 최소값(n_{ij})과 최대값(x_{ij})를 구하는 단계(18단계)와, 디스토션 확률밀도(Distortion density)를 구

하는 단계 (20단계)로 이루어진다.

즉, 상태세그먼테이션을 구하는 단계(10단계)는 학습을 위한 총 K개의 관찰 시퀀스들을 비터비(Viterbi) 알고리즘에 의하여 세그먼테이션한 후 백트래킹 순차(Backtracking procedure)에 의해 각각의 상태 세그먼테이션 시퀀스를 얻는다. 상기 제1도의 (a)는 비터비 알고리즘에 의해 학습관찰 시퀀스가 각각 상태 세그먼테이션 된 결과를 보여주고 있다. 특별히, 3번째 관찰 시퀀스에서 상태 2의 길이가 0인 경우를 보여주고 있다(즉 d_{23}=0). 실제로 비터비 절차 수행 후의 결과는 상기 제1도의 (b)와 같은 상태 시퀀스로 나타내어진다. 이에 대한 상태 지속(state duration)은 상기 제1도의 (c)와 같이 나타내어진다. 예를 들어 3번째 관찰 시퀀스에서 상태 1의 길이는 3으로서 d_{13}=3 으로 나타내고 있다.

전체 관찰시퀀스에 대한 각 상태(i)의 평균길이(mi)를 (식 1-1)에 의해 구하는 단계(12단계)는 다음과 같다.

$$m_i = \sum_{k=1}^{K} d_{ik} \Big/ \sum_{k=1}^{K} 1, \quad 1 \le i \le N \qquad \text{(식 1-1)}$$

$\sum_{k=1}^{K} 1$: 상태(i)의 duration 길이가 0 아닌 학습 관찰 시퀀스 개수 상기 제1도의 (d)에 도시된 바와 같이 관찰 시퀀스 전체에 대해 상태 세그먼테이션을 하고, 각 상태에 대한 평균길이를 구하는 것이다. 예를 들어, 상태 1의 평균 길이는 3, 상태 2의 평균길이는 2이다. 상태 평균길이(mi)를 고려하여 각각의 학습 관찰 시퀀스를 상태별로 시간구간을 정규화한다. 1번째 관찰 시퀀스 상태 1의 경우 길이가 4로 세그먼테이션 되었지만, 정규화에 의해 3등분 되는 시간 구간만 고려된다.

또한 상기 (식 1-1)에서 분모를 k로 하는 것이 평균 길이를 구한다는 관점에서 보면 더 적절하다고도 할 수 있으나, 디스토션의 확률밀도(density)를 구할 때, 상태 지속이 없는 경우(즉, d_{ik}=0 인 경우)의 확률밀도는 따로 취급했으므로 상기 (식 1-1)과 같이 정의하는 것이 오히려 적당하다고 할 수 있다.

상기 제10단계의 각 상태(i)에서의 정규화 시간구간(j)마다 평균 벡터(v_{ij})를 (식 1-2)와 (식 1-3)에 의해 구하는 단계(14단계)는 다음과 같다.

for $i = 1, 2, \ldots, N$, for $j = 1, 2, \ldots, m_i$

$$t = d_{ik} \frac{j}{m_i} = \sum_{l=1}^{j} d_{il} \qquad \text{(식 1-2)}$$

$$v_{ij}(l) = \sum_{k=1}^{K} o_{ik}(t) \Big/ \sum_{k=1}^{K} 1, \quad 1 \le l \le L \qquad \text{(식 1-3)}$$

상기 (식 1-2)는 상태(i)의 j번째 정규화 시간 구간의 평균벡터를 구하기 위하여 k번째 관찰 시퀀스에서 계산될 시간 구간의 인덱스(t)를 결정하기 위한 식으로서, 앞선 상태의 평균 시간 구간 수(즉 상태 길이)를 모두 더한 뒤, 상태(i)에서의 정규화하여 j번째인 시간 구간의 인덱스를 합한 것이다.

상기 제1도의 (e)는 평균벡터(v_{ij})를 보여 주고 있다. 평균벡터에서의 상태 당 시간 구간 수는 상기 12단계에서 구해진 길이를 기준으로 각각의 학습 관찰 시퀀스를 상태별로 정규화하여 각 정규화 시간 구간

$d_{11}(=4) \times \frac{1}{m_1}(=3)$

에 대한 평균 벡터를 구한다. 예를 들어 v_{12}은, 관찰 시퀀스 1의 번째 시간 구

$d_{12}(=2) \times \frac{1}{m_1}(=2)$

간, 관찰 시퀀스 2의 번째 시간 구간, ... 관찰 시퀀스 k의

$d_{1k}(=4) \times \frac{1}{m_1}(=4)$

번째의 시간 구간의 총합을 상태 1의 지속이 0이 아닌 관찰 시퀀스 개수로 나눔으로써 구해진다.

각각의 학습관찰 시퀀스(k)에 대해 각 상태(i)에서의 정규화 시간 구간(j) 마다 평균 벡터에 대한 디스

토션(eijk)을 다음의 (식 1-4)와 (식 1-5)를 통해 구하는 단계(16단계)는 다음과 같다.

for i = 1,2,3,....N , for j=1,2,3,....mi

$$d_{ij} = \sum_{i}^{j} d_{ij} \qquad \text{(식 1-4)}$$

$$e_{ijk} = distance(r_{ij}, d_{ik}) \quad , \quad d_{ik} \neq 0 \qquad \text{(식 1-5)}$$

상기 제2도의 (b)는 관찰 시퀀스의 상태별 시간 구간마다 구해진 디스토션을 나타내고 있다. 예를 들어 e_{123}은 3번째 관찰 시퀀스의 상태 1의 2번째 정규화 시간 구간에 해당하는 관찰벡터의 평균벡터 U_{12}에 대한 거리(distortion)를 나타낸다. 상기 (식1-5)의 distance는 LPC-캡스트럼 계수의 경우는 유클리디언 거리를 사용할 수 있다. 하지만 실제로 시스템 구현 시에는 유클리디언 거리의 계산식에 포함되어 있는 제곱근과 같은 연산은 수행하지 않아도 된다. 디스토션 확률밀도를 구할 때에는 시간 구간 사이의 상대적인 거리(Distortion)만이 필요하기 때문이다.

각 상태(i)에서의 정규화 시간구간(j)마다 총 k개의 학습 관찰벡터들의 디스토션 중 아래(식 1-6)와 (식 1-7)에 의해 최소값(h_{ij})과 최대값(x_{ij})을 구하는 단계(18단계)는 다음과 같다.

$$x_{ij} = \min(e_{ijk}) \quad , \quad 1 \le k \le K \quad , \quad 1 \le j \le m_i \quad , \quad 1 \le i \le N \qquad \text{(식 1-6)}$$

$$y_{ij} = MAX(e_{ijk}) \quad , \quad 1 \le k \le K \quad , \quad 1 \le j \le m_i \quad , \quad 1 \le i \le N \qquad \text{(식 1-7)}$$

상기 제2도의 (c)는 상기 18단계의 결과를 나타내고 있다. 전체 K개의 학습관찰 시퀀스에 대하여 각 정규화 시간 구간마다 거리(Distortion)의 최대값과 최소값을 구한다.

디스토션 확률밀도(Distortion density)를 아래의 (식1-8), (식1-9) 및 (식 1-10)을 통해 구하는 단계(20단계)는 다음과 같다.

for i = 1,2,3, ... , N , for j = 1,2,3, ... m_i

$p_{ij}(0)$ = (d_{ij} 보다 작은 학습관찰 시퀀스 k의 갯수)$\frac{1}{K}$, $s = 0$

$$p_{ij}(s) = \begin{cases} (e_{ijk} \text{ 중 } n_{ij} + \left[\frac{(s-1)(x_{ij}-n_{ij})}{S}\right] - \delta \text{ (if only } s=1), \\ \quad s\frac{(x_{ij}-n_{ij})}{S} + \delta \text{ (if only } s=S) \text{ 인} \\ \text{학습 관찰 시퀀스 } K\text{의 갯수})/K \quad , 1 \le s \le S \end{cases}$$

$p_{ij}(OUT) = 0$, otherwise

상기 제2도의 (d)에 도시된 바와 같이 디스토션 확률 밀도(Distortion density)는 정규화 시간 구간(normalized frame)마다 존재한다. 상기 디스토션 확률밀도(Distortion density)는 거리(distortion)의 최소값과 최대값 사이를 S개의 구간으로 나누어 각 시간 구간의 관찰 벡터의 거리(distortion)가 상기 거리(distortion) 구간(s)에 속하는 개수(학습 관찰 시퀀스의)를 이용하여 확률밀도(density)를 형성한다.

예를 들어, $p_{11}(2)$는 상태 1의 1번째 정규화 시간 구간에서 distortion 구간 2가 차지하는 확률을 나타내는 것으로서, e_{111}부터 e_{11k}까지의 distortion 중에서 구간 2에 속하는 개수를 전체(K)로 나눈 값이다.

양끝이 구간 0과 S는 δ값에 의해 조금씩 넓어진다. 학습과정에서 디스토션 확률밀도(density)를 구성하는 중에서는 δ의 값이 의미가 없으나(최소값과 최대값을 벗어나는 거리는 존재하지 않으므로), 인식시에는 인식을 위한 입력 시퀀스의 시간 구간 거리의 양끝의 구간에서 벗어나는 양이 δ보다 작을 때에는 구간 0과 S에 포함되므로 어떤 의미를 가지게 된다.

제4도는 HMM/Distortion 인식기의 인식과정의 블럭 다이어그램이다.

인식단계에서 사용되는 데이타는 벡터 양자화에 의한 코드북(41)과 일반적인 HMM 에 의해 구해진 모델파라미터(43), 및 거리 확률밀도(450이다.

인식하고자 하는 데이타가 입력되면 이는 코드북에 의해 적절한 이산적인 코드북 인덱스(42)로 이루어진 관찰 시퀀스로 변화되며, Viterbi 절차(44)에 의해 $\rho(o \setminus \lambda)$를 구한다. 또한 앞서 수행된 Viterbi 절차(44)의 부수적인 결과인 세그먼테이션된 상태 시퀀스(46)와 입력 관찰 벡터(LPC-켑스트럼)의 디스토션으로부터 디스토션 확률을 구하여 비터비 알로리즘의 log-likelihood score를 증가시켜 인식한다.

여기서, 인식과정에서 디스토션 확률을 결합하기 위한 방법은 다음의 3단계로 이루어진다.

첫 번째 단계에는 Viterbi 알고리즘으로 입력 관찰 시퀀스를 상태 세그먼테이션하여 각 상태의 지속시간(d_i)를 구하고, 두 번째 단계에서는 상기 각 상태(i)에서의 각 정규화 시간 구간(j)마다 학습 중에 구해진 평균벡터(υ_{ij})와의 거리를 구하여 e_{ij}가 속하는 구간(s)을 결정한 후 구간을 p_{ij}에 대입하여 아래의 식(1-11)에서 거리 확률(p_{ij}(s))을 구한다.

$$\text{for } i = 1,2,3, \ldots, N \qquad \text{for } j = 1,2,3, \ldots, m_i$$

$$t = d_i \,/\, m_i \quad ; \quad \sum_{i=1}^{N} d_i$$

$$c_{ij} = distance\ (\ c_{ij}, \ldots, c_{ij}\ )$$

$$s = c_{ij} \text{가 속해 있는 } distance \text{ 구간 , if } d_i \neq 0$$

$$= 0 \qquad \text{, if } d_i = 0$$

$$= out \qquad \text{, } c_{ij} \notin [\ n_{ij} - \delta\ ,\ x_{ij} + \delta\ ]$$

각 시간 구간의 거리가 s에 속하는지를 검사하기 때문에 구간 1과 S의 범위를 넓히는 δ값의 설정은 인식률에 영향을 미치게 된다. 또한 d_i가 0인 경우, 즉 입력 관찰 시퀀스의 상태(i)에서의 길이가 0인 경우에는 구간 0에 속하게 된다.

세 번째 단계는 인식확률의 로그-라이크후드 스코어(log-likelihood score)를 다음 식과 같이 결정한다.

$$\log P(q,\ o \mid \lambda) = \log P(q,\ o \mid \lambda) + \alpha \sum_{i=1}^{N} \sum_{j=1}^{M} \log[p_{ij}(s)]$$

상기 α는 distortion 확률에 곱해지는 scaling factor로서, 확률계산을 위한 log P(q, $o \setminus \lambda$)의 크기와 맞추기 위해 scaling하는 정도와 전체 log likelihood score에 대한 distortion확률이 차지할 수 있는 비중에 의해서 경험적으로 결정된다.

상술한 바와 같은 본 발명에 의하면, 학습과정에서, 일반적인 HMM에 의해 학습한 후, 비터비(Viterbi) 알고리즘에 의해 구할 수 있는 각 학습 데이타의 최적 상태 천이(maximum likelihood state transition)를 형성하여 각 상태에서 생성되는 관찰 심볼에 대한 디스토션 확률밀도를 구하여 이를 인식기에 이용하기 때문에 관찰 시퀀스에 대한 시간적인 구조를 고려할 수 있으며, 상태 세그먼테이션(state segmentation)이후 상태별로 소속된 정규화 시간 구간마다 관찰 벡터(즉LPC-캡스트럼)에 대한 디스토션 확률밀도를 구하여 인식시에 사용하므로 관찰시퀀스의 시간적인 구조와 학습데이타의 전처리 수준의 특징(LPC-캡스트럼의 특성)을 동시에 고려할 수 있는 효과를 제공한다.

(57) 청구의 범위

청구항 1

디스토션 확률밀도를 가진 은닉마코프모델(HMM/Distortion)을 이용한 음성인식방법에 있어서, 학습데이타를 벡터양자화시켜 코드북을 만들고, 상기 벡터양자화된 학습데이타의 모델파라메터를 재추정한 후, 비터비 알고리즘에 의해 상태시퀀스를 세그먼트하여 모델파라메터를 생성하고, 상기 세그먼트된 상태시퀀스를 근거로 상기 학습데이타의 디스토션 확률밀도를 추정하는 학습과정과, 상기 학습과정에서 생성된 코드북과 모델파라미터 및 디스토션 확률 밀도를 이용하여, 인식하고자 하는 데이타가 입력되면, 상기 코드북을 이용하여 상기 입력데이타를 관찰시퀀스로 변환시켜, 비터비 알고리즘에 의해 구해진 확률($\rho(o \setminus \lambda)$)과, 상기 비터비 절차에 의해 생성된 세그먼트된 상태시퀀스와 입력 관찰 벡터(LPC-켑스트럼)의 디스토션으로부터 구해진 디스토션 확률을 연산하여 음성을 인식하는 과정을 포함하는 것을 특징으로 하는 음성인식방법.

청구항 2

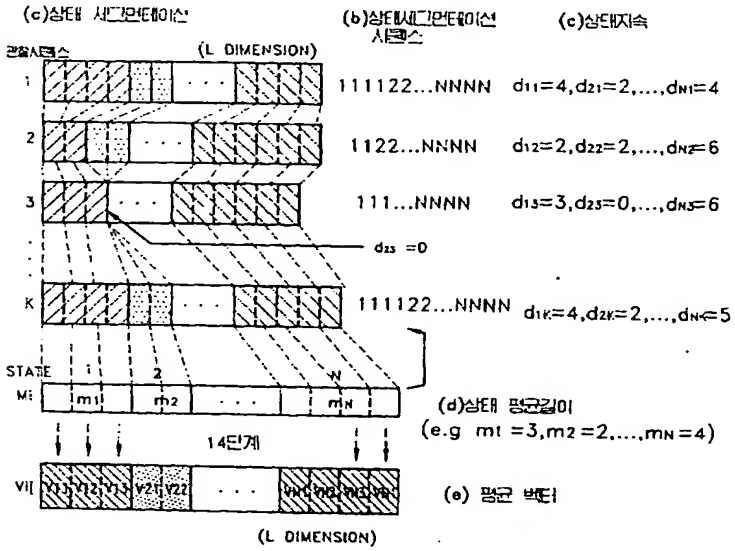
제1항에의 상기 학습과정에서, 상기 디스토션 확률밀도를 구하는 방법은 상태 세그먼테이션 시퀀스를 구하는 단계; 상기 상태 세그멘테이션된 학습관찰 시퀀스를 참조하여 각 상태(i)의 평균길이(mi)를 구하는 단계; 상기 상태(i)에서의 정규화 시간 구간(j) 마다 평균 벡터(υ_{ij})를 구하는 단계; 각각의 학습관찰 시퀀스(k)에 대한 디스토션(eijk)을 구하는 단계; 및 각 상태(i)에서의 정규화 시간구간(j) 마다 총 k개의 학습 관찰 벡터들의 디스토션 중 최소값(n_{ij})과 최대값(x_{ij})를 구하는 단계를 포함하는 것을 특징으로 하는 음성인식방법.

청구항 3

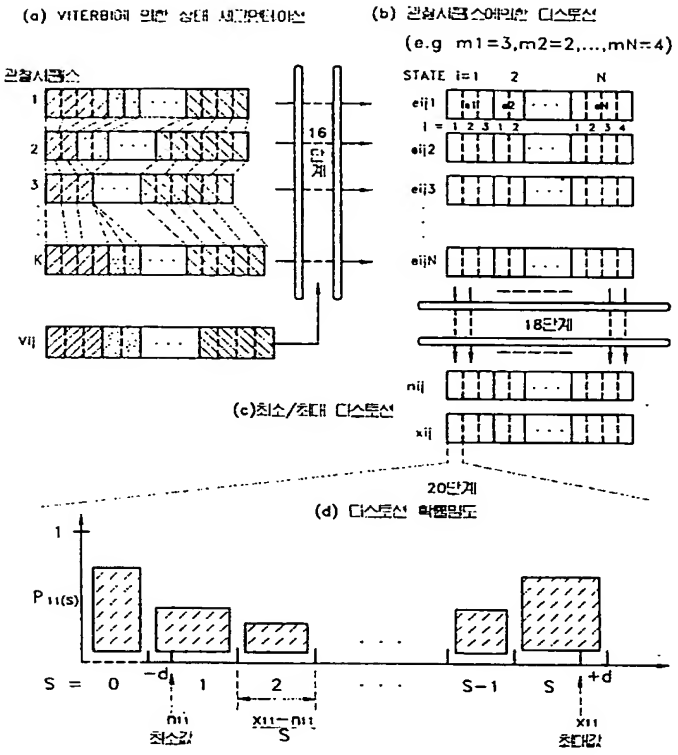
제1항의 상기 인식과정에서, 디스토션 확률밀도를 결합시키는 방법은 비터비 알고리즘으로 입력 관찰 시퀀스를 상태 세그먼테이션하여 각 상태 지속(d_i)를 구하는 단계; 상기 각 상태(i)에서의 각 정규화 시간구간(j)마다 학습 중에 구해진 평균벡터(υ_{ij})와의 거리를 구하여 e_{ij}가 속하는 구간(s)을 결정한 후 이 구간(s)을 ρ_{ij}에 대입하여 거리 확률($\rho_{ij}(s)$)을 구하는 단계; 및 인식확률의 로그-라이크후드 스코어(log-likehood score)를 결정하는 단계를 포함하는 것을 특징으로 하는 음성인식방법.

도면

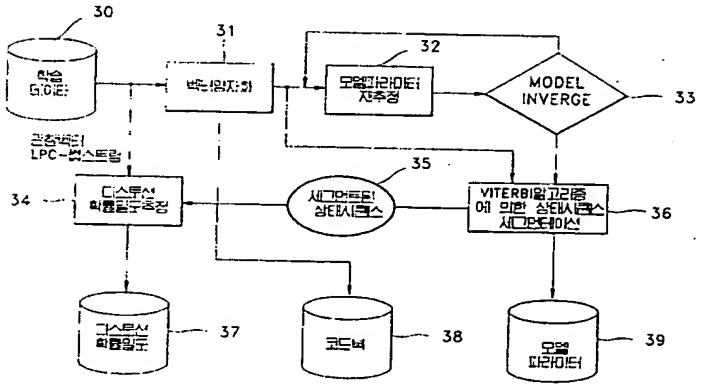
도면1



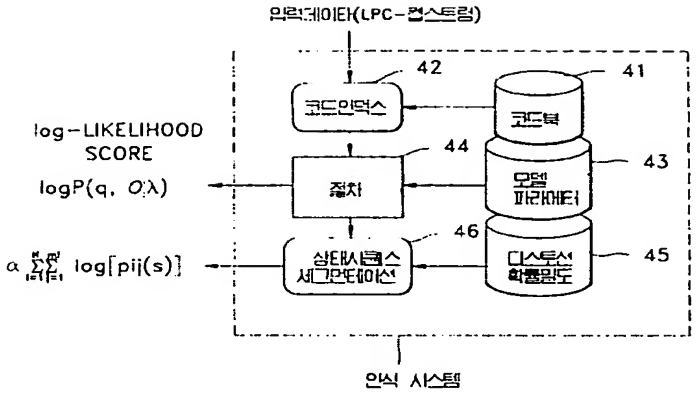
도면2


(a) VITERBI에 의한 상태 세그먼테이션
(b) 관찰시퀀스에의한 디스토션
(e.g m1=3,m2=2,...,mN=4)
관찰시퀀스
STATE i=1 2 N
16단계
18단계
(c)최소/최대 디스토션
20단계
(d) 디스토션 확률밀도
최소값
최대값

도면3


30
31
32
MODEL INVERGE
33
34
35
36
37
38
39

도면4


입력데이터(LPC-켑스트럼)
42
41
코드인덱스
코드북
log-LIKELIHOOD
SCORE
44
43
logP(q, O|λ)
정차
모델
파라메터
46
45
상태시퀀스
세그먼테이션
디스토션
확률밀도
인식 시스템